NUCLEAR SCIENCE AND TECHNIQUES **26**, 030602 (2015)

GA-based dynamical correction of dispersion coefficients in Lagrangian puff model*

MA Yuan-Wei (马元巍), WANG De-Zhong (王德忠),† and JI Zhi-Long (吉志龙)
School of Nuclear Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
(Received June 19, 2014; accepted in revised form September 20, 2014; published online June 20, 2015)

In atmospheric dispersion models of nuclear accident, the dispersion coefficients were usually obtained by tracer experiment, which are constant in different atmospheric stability classifications. In fact, the atmospheric wind field is complex and unstable. The dispersion coefficients change even in the same atmospheric stability, hence the great errors brought in. According to the regulation, the air concentration of nuclides around nuclear power plant should be monitored during an accident. The monitoring data can be used to correct dispersion coefficients dynamically. The error can be minimized by correcting the coefficients. This reverse problem is nonlinear and sensitive to initial value. The property of searching the optimal solution of Genetic Algorithm (GA) is suitable for complex high-dimensional situation. In this paper, coupling with Lagrange dispersion model, GA is used to estimate the coefficients. The simulation results show that GA scheme performs well when the error is big. When the correcting process is used in the experiment data, the GA-estimated results are numerical instable. The success rate of estimation is 5% lower than the one without correction. Taking into account the continuity of the dispersion coefficient, Savitzky-Golay filter is used to smooth the estimated parameters. The success rate of estimation increases to 75.86%. This method can improve the accuracy of atmospheric dispersion simulation.

Keywords: Adaptive parameter, Genetic algorithm, Atmospheric dispersion, Nuclear accident

DOI: 10.13538/j.1001-8042/nst.26.030602

I. INTRODUCTION

Traditionally, coefficients of radionuclide dispersion were preset for nuclear accident. They are obtained from field campaign experiments, with the differences between meteorological conditions in experiments and real-time meteorological conditions being a major source of errors. In most dispersion models, Pasquill-Gifford (*PG*) curve established on Prairie-Grass-Field experiment [1] and similar dispersion coefficients system are used to represent the lateral and vertical plume spread rates.

Generally these dispersion coefficient systems are used in conservative forecast and recommended by regulation. But when the coefficients are validated by tracer experiments in nuclear power plants (NPPs), the result showed that the dispersion coefficients are much different with the ones fitted by the experimental data [2]. In an NPP, air concentration of radionuclides should be monitored during an accident [3]. For more accurate dispersion coefficients, a viable way is to correct the empirical ones with monitored data dynamically. The dispersion coefficients will be corrected only based on monitored data and will not be calculated by atmospheric stability and other information.

Haupt proposed to coupled the genetic algorithm (GA) with Gaussian plume model and receptor model, so as to achieve a better estimation of pollutant's dispersion [4]. Jeong used a similar technique to analyze the atmospheric dispersion of nuclides from an NPP [5, 6]. Allen used the assimilation estimation method to estimate wind-field information in the dispersion model [7]. Most numerical simulations used ideal models. However, the correlation coefficient value between predicted values and observed values is less than 0.5, and value of observation has a significant noise [8]. Thus in the actual situation, this error should not be ignored. The above studies indicate that the source term estimation and dispersion parameter dynamic correcting are effective methods to improve the accuracy of dispersion models. Nonetheless, the impact of dispersion coefficient error on the result requires further study.

To find out the performances with diverse errors of dispersion coefficient, a scheme was established based on GA method and was validated by numerical simulation and Kincaid experiment data [8].

II. METHOD

A. Lagrangian puff model

Parameters estimation requires that the dispersion model must be suitable for complex meteorological conditions, and high computing efficiency. Compared to Gaussian plume model and particle model, Lagrangian puff model meets the requirements better [9]. Therefore, it was used as a forward model to calculate the pollutant distributions.

The basic assumption of Lagrangian puff model is that sequentially releasing of pollutant is treated as a series of Gaussian shaped puffs at a fixed rate. The concentration of pollutants in each puff fits Gaussian distribution. The total concentration at some location is the sum of all puffs' value at this point. The concentration distribution of a puff can be described as Eq. (1)

* Supported by the National Natural Science Foundation of China (No. 11175118) and Science and Innovation Project of Shanghai Education Commission (No. 12ZZ022)
† Corresponding author, dzwang@sjtu.edu.cn

$$C(x,y,z)=\frac{Q}{\left(\sqrt{2\pi}\right)^3\sigma_x\sigma_y\sigma_z}\cdot \exp\left[-\frac{1}{2}\left(\left(\frac{x-x_{\rm c}}{\sigma_x}\right)^2+\left(\frac{y-y_{\rm c}}{\sigma_y}\right)^2\right)\right]\cdot \left[\exp\left(-\frac{1}{2}\left(\frac{z-z_{\rm c}}{\sigma_z}\right)^2\right)+\exp\left(-\frac{1}{2}\left(\frac{2z_{\rm inv}-z_{\rm c}}{\sigma_z}\right)^2\right)\right], \tag{1}$$

where Q is total concentration of the puff; ($x_{\rm c}$, $y_{\rm c}$, $z_{\rm c}$) are coordinates of the puff center; $z_{\rm inv}$ is the height of inversion temperature layer; (x, y, z) are coordinates of the observation point; and (σ_x, σ_y, σ_z) are dispersion parameters of the puff.

There are many empirical formulas to compute the dispersion parameters according to the atmospheric stability and downwind distance, among which the PG curve, as described in Eq. (2), is most often used.

$$\begin{aligned}\sigma_x &= \sigma_y,\\ \sigma_y &= p_y\cdot x^{q_y},\\ \sigma_z &= p_z\cdot x^{q_z}.\end{aligned} \tag{2}$$

The σ_x, σ_y and σ_z are computed according to downwind distance x, and the PG coefficients p_y, q_y and p_z and q_z, are determined by the wind-field and atmospheric stability.

B. *PG* coefficients correction

PG curve is derived from Prairie Grass field experiments. If the experimental conditions differ from the actual meteorological conditions, errors will be introduced to the model. To minimize effects of the PG error, one way is using the original PG coefficients as initial values, and correcting them by the observed values. This process can be regarded as an optimization by minimizing the difference between calculated values and observed values.

C. Genetic algorithm

The reverse model of atmospheric dispersion is strongly nonlinear and sensitive to initial values. Tan used the least square method to estimate the dispersion coefficients but most of the results are local optimal solutions [10]. This means that an effective global optimization algorithm is required, and GA can just satisfy this requirement. Invented by J. Holland in 1975, GA is an efficient and parallel global search method.

Fitness function (ft) is an essential issue in GA. It determines differences between the observed and predicted value. The smaller the ft is, the closer between observations and predictions. On the other hand, an observation's weight is different from each other due to the diversity of observing conditions. An appropriate fitness function should comply with two principles: 1) for observation station i, the fitness function value should increased with the differences between the observed value O_i and calculated value C_i; 2) observation stations should be given different weight in the fitness function due to their data convince.

Ji *et al.* compared four types of fitness function, and found that concentration-related weight of observations is better than equal weight [11]. The fitness function can be constructed as Eq. (3) on basis of the least square method

$$ft=\frac{\sum\limits_{i=1}^{N}O_i(C_i-O_i)^2}{\sum\limits_{i=1}^{N}O_i}, \tag{3}$$

where, N is the number of observation stations; C_i is the predicted value in Station i; and O_i is the observation value in Station i. This fitness function is a modification for equal weight of every station. The weight is related to the concentration value of observations.

III. NUMERICAL SIMULATIONS

The numerical stability of this estimation scheme was validated by numerical simulations. White noise (a sequence of serially uncorrelated random variables with zero mean and finite variance) was added to the observations which were calculated by the model. Based on modified observations, GA was used to estimate the PG coefficients. Performance of the scheme was evaluated by comparing the 'true' PG coefficients with the estimated ones.

A. Setup of dispersion parameters

To simplify the calculation, in a Cartesian coordinate system, the release point was set to (0, 0), height is 187 m, release rate was one unit, and tracer gas was SF_6. Wind was stable with one direction. Atmospheric stability was D class. The original PG coefficients, $PG^{\rm true}$, were set as Eq. (4). The changing of PG coefficients were estimated based on the observation value in each simulation step.

$$\begin{aligned} p_y^{\rm true} &= 1.503, & q_y^{\rm true} &= 0.833,\\ p_z^{\rm true} &= 0.151, & q_z^{\rm true} &= 1.219.\end{aligned} \tag{4}$$

B. Configuration of observations

Signal-to-noise ratio (*SNR*) is defined as the ratio of signal power to the noise power:

$$SNR=\sigma_{\rm signal}/\sigma_{\rm noise}. \tag{5}$$

Different *SNR* levels of 50, 25, 5, 2.5 and 0.25 were added to every sampling, including positive infinity. The noise is

white noise, which is a random signal with a constant power spectral density. According to the modified observations, the PG^{corr} was estimated to analyze the impact of the error.

Initial values of parameters were set as Eq. (6), with every parameter added some noise:

$$\begin{aligned} p_y^{\mathrm{pre}} &= 1.533, \quad q_y^{\mathrm{pre}} = 0.712, \\ p_z^{\mathrm{pre}} &= 0.255, \quad q_z^{\mathrm{pre}} = 1.401. \end{aligned} \tag{6}$$

GA with the fitness function corrected these PG^{pre} to achieve PG^{corr}. Less differences between PG^{corr} and PG^{true} means better performance of the fitness function.

C. Result of numerical experiment

The error between PG^{corr} and PG^{true} is expressed by Eq. (7)

$$\sigma_{pq} = \sqrt{(\log_2 p_y^{\mathrm{corr}} - \log_2 p_y^{\mathrm{true}})^2 + (q_y^{\mathrm{corr}} - q_y^{\mathrm{true}})^2 + (\log_2 p_z^{\mathrm{corr}} - \log_2 p_z^{\mathrm{true}})^2 + (q_z^{\mathrm{corr}} - q_z^{\mathrm{true}})^2}. \tag{7}$$

As shown in Fig. 1, the error of estimation parameters increased with $1/SNR$. The variance of white noise is $1/SNR$, and at $1/SNR = 1$, $\sigma_{pq} < 1.5$, which means the PG^{corr} is very close to PG^{true}. It indicates that this scheme has good robustness.

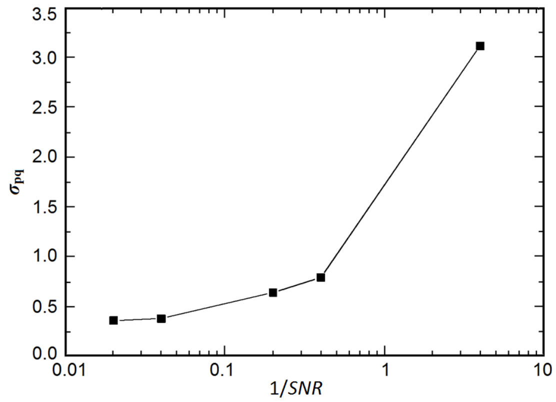

Fig. 1. Correction result with diverse errors.

IV. EXPERIMENT VALIDATION

The simulation shows that parameters of the dispersion model can be corrected on-line by GA. But in a tracer experiment, conditions differ from the simulation. Being not the ideal white noise, the errors distribute disorderly, and are larger and time-dependent. So, Kincaid tracer experiment data were used to validate the effectiveness of GA.

A. Kincaid tracer experiment

The Kincaid field experiment was performed as part of the EPRI Plume Model Validation and Development Project. A very comprehensive experimental campaign was conducted in 1980 and 1981. The Kincaid NPP is situated in Illinois, USA (39.59°N, 89.49°W). The terrain is at an elevation of ∼180 m, thus in the model the terrain effect is not considered.

The NPP has a 187 m stack of Φ9 m. In the experiment, SF_6 was released from the stack. The tracer releases started several hours before the sampling. There were ∼350 h of tracer experiments in the experimental campaign.

The 1284 arc-max concentration data, of "high quality" and classified artificial, are usually used for model validation and development. However, for this estimation, data should not be screened. In a real accident, there is not enough information to judge the effectiveness of data. All the data was used, which introduced uncertainty to the estimation.

B. Validation method

There were 20 cases of experiment, each lasted 3 h to 9 h. Over 100 observation stations recorded the concentration of SF_6 in the air during the release period. The statistical values, *FA2, FA5, FB, NMSE, CORR* and *BIAS*, as described as below, are used to measure the differences between predicted and observed values.

$$\begin{aligned} &FA2 = \text{fraction of data satisfying} \quad 0.5 < C_\mathrm{o}/C_\mathrm{p} < 2, \\ &FA5 = \text{fraction of data satisfying} \quad 0.2 < C_\mathrm{o}/C_\mathrm{p} < 5, \\ &FB = (\bar{C}_\mathrm{o} - \bar{C}_\mathrm{p})/[0.5(\bar{C}_\mathrm{o} + \bar{C}_\mathrm{p})], \\ &NMSE = \overline{(C_\mathrm{o} - C_\mathrm{p})^2}/(\bar{C}_\mathrm{o}\bar{C}_\mathrm{p}), \\ &CORR = \overline{(C_\mathrm{o} - \bar{C}_\mathrm{o})(C_\mathrm{p} - \bar{C}_\mathrm{p})}/(\sigma_\mathrm{o}\sigma_\mathrm{p}), \\ &BIAS = \bar{C}_\mathrm{o} - \bar{C}_\mathrm{p}, \end{aligned} \tag{8}$$

where, p denotes model prediction; o denotes observation; over bar denotes averaged dataset and σ is the mean square error over the dataset.

V. RESULTS AND DISCUSSION

A. Results of GA estimation

The results are shown in Table 1. In the 'Original data' row, the correlation coefficient (*CORR*) is 0.20, indicating a weak positive relationship between the data of prediction and observation; the *FA*2(0.11) and *FA*5(0.26) mean that the differences

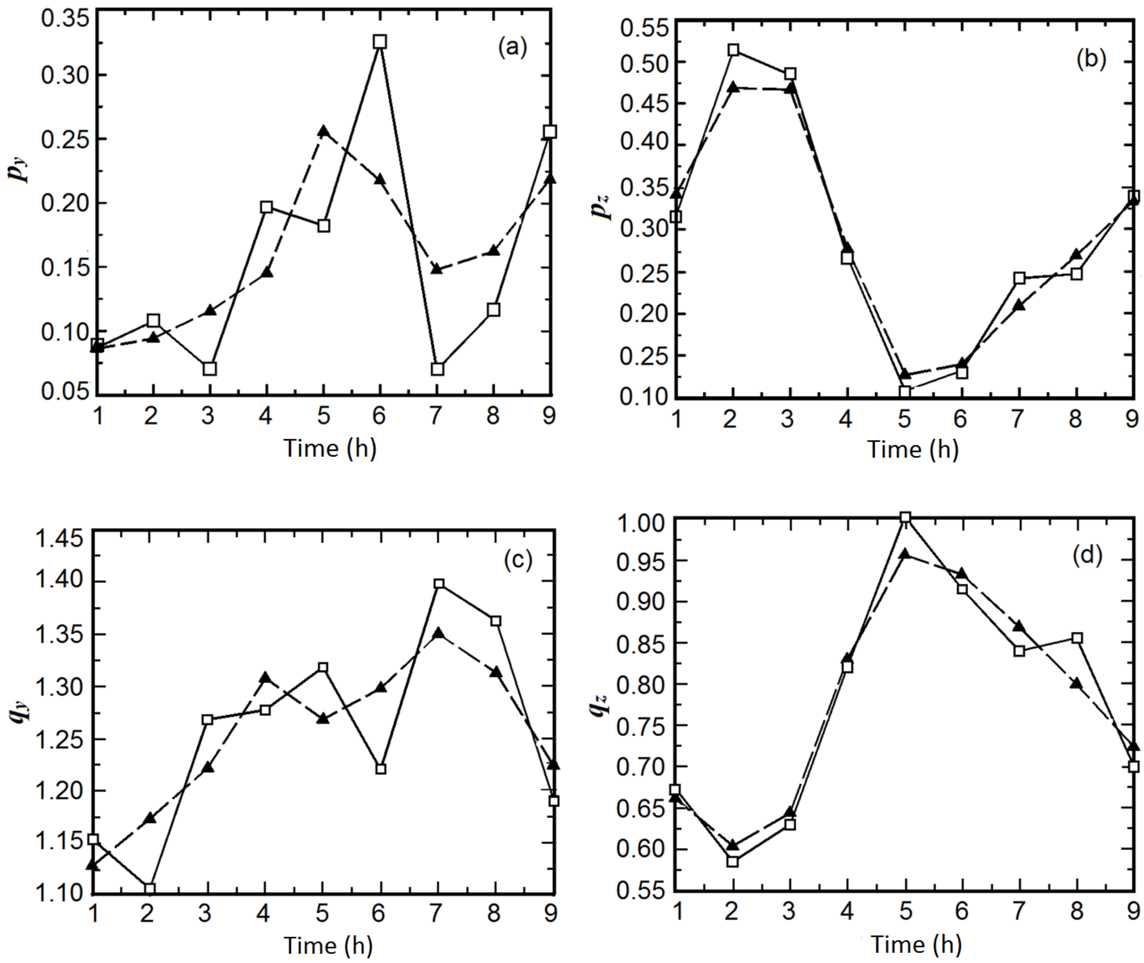

Fig. 2. Smoothing filter of estimated PG coefficients (a) p_y, (b) p_z (c) q_y and (d) q_z (□: original data; ▲: smoothed data).

TABLE 1. Influence of PG correction on forecast ability of dispersion model

	NMSE	*BIAS*	*CORR*	*FA*2	*FA*5	*FB*	Success rate (%)
Original data	24.61	−0.87	0.20	0.11	0.26	−1.05	75.86
GA	22.97	−0.13	−0.03	0.18	0.37	0.43	69.54
GA Prediction	25.57	−0.07	−0.02	0.15	0.34	0.47	70.11
GA smooth	31.19	−0.02	−0.01	0.15	0.32	0.65	75.86

are big. In view of this, estimating PG coefficients is necessary to improve the model quality. Using GA method (the 'GA' row) increases the *FA* value and decreases the *BIAS* and *FB* values. Therefore, after the dynamic correction, the model is more accurate and efficient with the large errors of dispersion coefficients.

B. GA Prediction

The prediction process uses the parameters that are fitted based on the historical data. For every case, the steps are as follows:

1) With the standard PG parameters as prior value, correct the parameters using GA in every step.

2) Calculate the prediction value based on the last step of the corrected parameters.

3) Calculate the prediction value based on the standard PG parameters.

4) Compare the prediction observations between Steps 2 and 3, it will tell the effect of parameters predicted by GA.

The results are given in Table 1. Most of the data worsened because the calculation of dispersion parameters in one step was not related to the next step. The result obtained by GA is only numerically best, not physically best. The continuing property of the PG coefficients was ignored.

C. Smoothing filter

Though the statistics values are much better than the ones in original model, the GA-estimated results are numerical instable. Some estimated values were greatly deviated from the true value. The success rate of estimations was 5% lower than the one without correction. To reduce impact of the

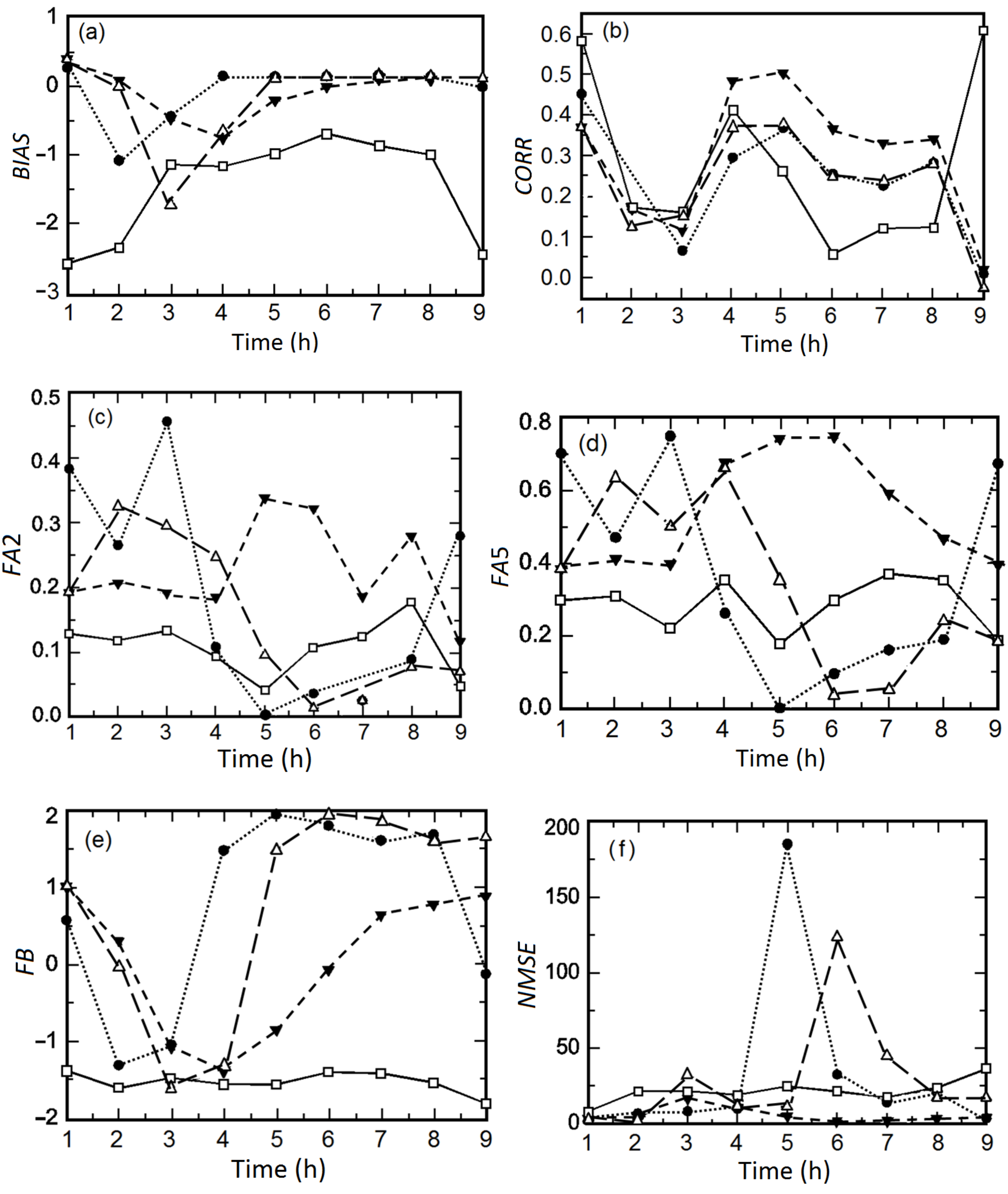

Fig. 3. Statistics values of (a) *BIAS*, (b) *CORR*, (c) *FA*2, (d) *FA*5, (e) *FB* and (f) *NMSE*, of case 50780 for estimation schemes of GA (●), GA Prediction (△) and GA Smooth (▼), comparing with the original data (□).

fluctuations in GA estimation, it assumed that the dispersion parameters were of continuous changes, as shown in Eq. (9)

$$\begin{aligned} p_y[t+1] &= p_y[t] + \eta, \\ q_y[t+1] &= q_y[t] + \eta, \\ p_z[t+1] &= p_z[t] + \eta, \\ q_z[t+1] &= q_z[t] + \eta, \end{aligned} \tag{9}$$

where η is white noise representing time-dependent changes of dispersion coefficients; t is the current time step and $t+1$ is the next time step.

Thus, before their uses for prediction, the GA-estimated parameters were filtered smoothly with Savitzky-Golay filter. The main advantage of this method is that it tends to preserve features of the distribution such as relative maxima,

TABLE 2. The correlation coefficient between FA5 and other statistics

	Step	Station No.	$\overline{C_p}/\overline{C_o}$	σ_p/σ_o	*NMSE*	*BIAS*	*CORR*	*FA*2	*FA*5	*FB*
GA	0.025	0.20[b]	0.32[b]	0.16[a]	−0.17[a]	−0.49[b]	0.22[b]	0.35[b]	0.32[b]	−0.43[b]
GA prediction	0.16[a]	0.093	0.23[b]	0.23[b]	−0.19[a]	−0.33[b]	0.12	0.17[a]	0.26[b]	−0.31[b]
GA smooth	0.13	0.13	0.25[b]	0.32[b]	−0.13	−0.47[b]	0.077	0.13	0.235	−0.43[b]

[a] Correlation is significant at the 0.05 level.
[b] Correlation is significant at the 0.01 level.

minima and width, which are usually 'flattened' by other adjacent averaging techniques. The information in the estimated dispersion coefficients will be reserved after Savitzky-Golay smoothing.

Take 50 780 case as example, the results after smoothing filter are shown in Fig. 2. The *PG* coefficients are more stable than before. It means that the distribution of nuclides changes smoothly step by step. Figure 3 shows statistics values of the results for the three schemes. 'GA smooth' scheme effectively improves the results, and inherited excellent characteristics of 'GA prediction' scheme. It balanced the stability and efficiency, performed best of the three schemes.

The statistical values of all the cases are shown in 'GA Smooth' row in Table 1. The *NMSE, FA2, FA5* and *FB* data are slightly worse than 'GA' and 'GA Prediction', but are still much better than those in the 'Original data'. The success rate is equal to that of 'GA', indicating that it removed the numerical instability. The 'GA Smooth' scheme estimates the dispersion coefficients efficiently.

D. Correlation test

To analyze which factor determined the accuracy of estimations, correlation tests were performed between value of *FA*5 and other statistics values. The results shown in Table 2 indicate that:

- *FA*5 has positive relationship with numbers of observation station, and with more observations, the prediction can be more accurate.
- The number of compute steps is not related to *FA*5, because the GA method does not take the historical data into consideration.

VI. CONCLUSION

Traditional dispersion coefficients are constant in the same class of atmospheric stability. A GA scheme was established to correct the dispersion coefficients of Lagrangian puff model dynamically. The correction coefficients could minimize the error. For big error of dispersion coefficients, the numerical simulation result shows that GA scheme performs well. After the correcting procedure of the Kincaid experiment data, the forecast ability of dispersion model is boosted significantly. Taking into account the continuity of the dispersion coefficient, Savitzky-Golay filter is used to smooth the estimated parameters. The success rate of estimation increases by 5%. The GA method coupled with Savitzky-Golay filter can improve the accuracy of atmospheric dispersion simulation efficiently.

[1] Davidson G. A modified power law representation of the Pasquill-Gifford dispersion coefficients. JAPCA J Air Waste Ma, 1990, **40**: 1146–1147. DOI: 10.1080/10473289.1990.10466761

[2] Leroy C, Maro D, Hébert D, *et al.* A study of the atmospheric dispersion of a high release of krypton-85 above a complex coastal terrain, comparison with the predictions of Gaussian models (Briggs, Doury, ADMS4). J Environ Radioactiv, 2010, **101**: 937–944. DOI: 10.1016/j.jenvrad.2010.06.011

[3] Xiang Y Y, Wang K, Zhang Y, *et al.* Radioactivity monitoring in environmental water and air around QNPP. Nucl Sci Tech, 2007, **18**: 316–320. DOI: 10.1016/S1001-8042(07)60069-4

[4] Haupt S E. A demonstration of coupled receptor/dispersion modeling with a genetic algorithm. Atmos Environ, 2005, **39**: 7181–7189. DOI: 10.1016/j.atmosenv.2005.08.027

[5] Jeong H J, Han M H, Hwang W T, *et al.* Application of data assimilation to improve the forecasting capability of an atmospheric dispersion model for a radioactive plume. Ann Nucl Energy, 2008, **35**: 838–844. DOI: 10.1016/j.anucene.2007.09.008

[6] Jeong H J, Kim E H, Suh K S, *et al.* Determination of the source rate released into the environment from a nuclear power plant. Radiat Prot Dosim, 2005, **113**: 308–313. DOI: 10.1093/rpd/nch460

[7] Allen C T, Young G S, Haupt S E. Improving pollutant source characterization by better estimating wind direction with a genetic algorithm. Atmos Environ, 2007, **41**: 2283–2289. DOI: 10.1016/j.atmosenv.2006.11.007

[8] Korsakissok I and Mallet V. Comparative study of Gaussian dispersion formulas within the polyphemus platform: evaluation with Prairie Grass and Kincaid experiments. J Appl Meteorol Clim, 2009, **48**: 2459–2473. DOI: 10.1175/2009JAMC2160.1

[9] Jung Y R, Park W G, Park O H. Pollution dispersion analysis using the puff model with numerical flow field data. Mech Res Commun, 2003, **30**: 277–286. DOI: 10.1016/S0093-6413(03)00024-7

[10] Tan W J, Wang D Z, Ma Y W, *et al*. Diffusion coefficient adaptive correction in Lagrangian puff model. Atom Energy Sci Technol, 2014, **48**: 571–576. DOI: 10.7538/yzk.2014.48.03.0571

[11] Ji Z L, Ma Y W, Wang D Z. Study on fitness functions of genetic algorithm for dynamically correcting nuclide atmospheric diffusion model. Nucl Tech, 2014, **37**: 040503. (in Chinese) DOI: 10.11889/j.0253-3219.2014.hjs.37.040503